| | |
|---|---|
| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |
| Ph. No. 9727753360 | Rajkot - 360001 |
| | Ph.No–(0281)2471645 |

## MBA SEMESTER 2

## Business Analytics- 4529201

## MODULE 3

**Data Mining:**

☐ Concepts and applications

- DM is an important part of Business Intelligence.

- Emergence in 1990s-2000s

- It is a field of computer science

- **Data mining** tools allow enterprises to predict future trends.

- Alternative Names:

    – Knowledge discovery(mining)

- Data Mining refers to extracting knowledge from large amount of data.

- The task of **discovering interesting patterns** from large amount of data.In great demand with wide applications.

Definition of DM

- The data mining is the process of analyzing the large amount of data which is stored in DW for useful information which makes the use of machine learning, artificial intelligence and statistical tools which may be undetected further.
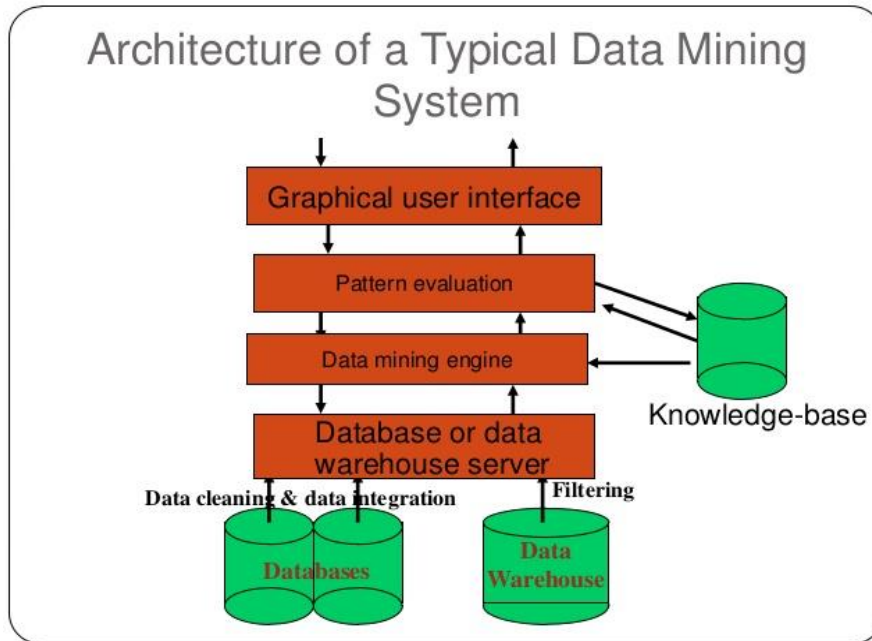
Architecture of DM

- **Data Base & Data Warehouse**

"a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process"

- **Data Cleansing, Data Selection & Data Integration**

  - **Data Cleansing :** Unnecessary data are removed.

  - **Data Selection  :** Relevant data is used for analytical purpose which are retrieved from DB.

  - **Data Integration :** Multiple Data sources may be combined.

- **Database or Data Warehouse Server :**

It can bring the data as per user requirement which is useful for DM task.

- **Data Mining Engine :**

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |

It performs various Data Mining operations such as associations , classification , Prediction & clustering

Ph.No–(0281)2471645

- **Pattern Evaluation:**

It is integrated one of the mining module which is 'KDD' and it helps to search only the useful pattern for evaluation.

- **Knowledge Base:**

This is used to search the useful pattern from hidden DB.

- **Graphical User Interface :**

This module is used to communicate between user & DM system and allow the user to browse DB from different Data Warehouse Schema.

Major DM tasks

- **Classifications** : Creation of pattern

- **Associations**: e.g. Along with 'A' , 'B' & 'C' occurs

- **Estimation**: predicting a continuous future

- **Deviation Detection** : finding changes

- **Link Analysis** : finding relationships

Application of DM

- Finance                        -        Credit Card Analysis

- Insurance              -        Claims Fraud Analysis

- Telecommunication    -     Call record analysis

- Transport                  -        Logistics management

- Consumer goods    -      Promotion analysis

- Scientific Research        -        Image, Video, Speech

- Utilities                          -        Power usage analysis

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

Lalpari Campus, Near Lalpari Lake,                    3 – Vaishalinagar

B/H Marketing Yard,                                  Nr. Amrapali Railway Crossing
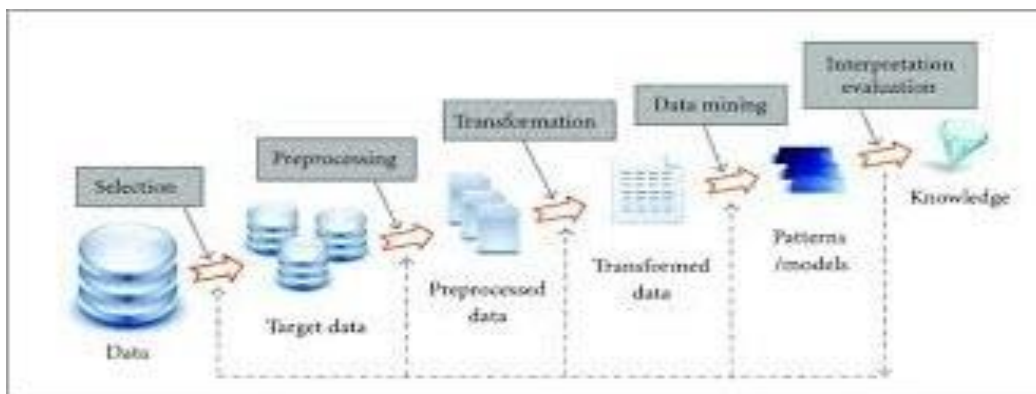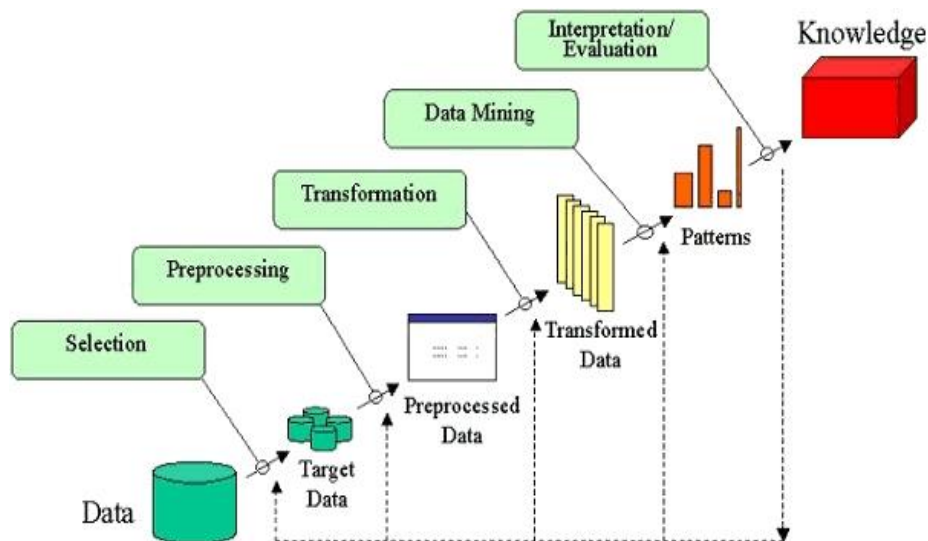
Amargadh (Bhichri), Rajkot                           Raiya Road, Rajkot

Rajkot - 360001

Ph. No. (0281) 2471645

- DM can be used in Public & Private Sector.

- It is used in major industries like banking, insurance, finance, retail, etc.

- Prediction model can be developed in mining to analyze and collect the data over a year.

- Effectiveness of a medicine can be predicted by the usage of mining.

- DM can be used by the researcher for disaster management & telecommunication purpose in the industry.

☐ Data mining process

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
|---|---|
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |
| Ph. No. 9727753360 | Rajkot - 360001 |

## #1) Data Cleaning

Data cleaning is the first step in data mining. It holds importance as dirty data if used directly in mining can cause confusion in procedures and produce inaccurate results.

Basically, this step involves the removal of noisy or incomplete data from the collection. Many methods that generally clean data by itself are available but they are not robust.

**This step carries out the routine cleaning work by:**

**(i) Fill The Missing Data:**

Missing data can be filled by methods such as:

- Ignoring the tuple.
- Filling the missing value manually.
- Use the measure of central tendency, median or
- Filling in the most probable value.

**(ii) Remove The Noisy Data:** Random error is called noisy data.

**Methods to remove noise are :**

**Binning:** Binning methods are applied by sorting values into buckets or bins. Smoothening is performed by consulting the neighboring values.

Binning is done by smoothing by bin i.e. each bin is replaced by the mean of the bin. Smoothing by a median, where each bin value is replaced by a bin median. Smoothing by bin boundaries i.e. The minimum and maximum values in the bin are bin boundaries and each bin value is replaced by the closest boundary value.

- Identifying the Outliers
- Resolving Inconsistencies

## #2) Data Integration

When multiple heterogeneous data sources such as databases, data cubes or files are combined for analysis, this process is called data integration. This can help in improving the accuracy and speed of the data mining process.

Different databases have different naming conventions of variables, by causing redundancies in the databases. Additional Data Cleaning can be performed to remove the redundancies and inconsistencies from the data integration without affecting the reliability of data.

Data Integration can be performed using Data Migration Tools such as Oracle Data Service Integrator and Microsoft SQL etc.

## #3) Data Reduction

This technique is applied to obtain relevant data for analysis from the collection of data. The size of the representation is much smaller in volume while maintaining integrity. Data Reduction is performed using methods such as Naive Bayes, Decision Trees, Neural network, etc.

**Some strategies of data reduction are:**

- **Dimensionality Reduction:** Reducing the number of attributes in the dataset.
- **Numerosity Reduction:** Replacing the original data volume by smaller forms of data representation.

- **Data Compression:** Compressed representation of the original data.

## #4) Data Transformation

In this process, data is transformed into a form suitable for the data mining process. Data is consolidated so that the mining process is more efficient and the patterns are easier to understand. Data Transformation involves Data Mapping and code generation process.

**Strategies for data transformation are:**

- **Smoothing:** Removing noise from data using clustering, regression techniques, etc.
- **Aggregation:** Summary operations are applied to data.
- **Normalization:** Scaling of data to fall within a smaller range.
- **Discretization:** Raw values of numeric data are replaced by intervals. **<u>For Example,</u>** Age.

## #5) Data Mining

Data Mining is a process to identify interesting patterns and knowledge from a large amount of data. In these steps, intelligent patterns are applied to extract the data patterns. The data is represented in the form of patterns and models are structured using classification and clustering techniques.

## #6) Pattern Evaluation

This step involves identifying interesting patterns representing the knowledge based on interestingness measures. Data summarization and visualization methods are used to make the data understandable by the user.

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| | |
|---|---|
| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |

## #7) Knowledge Representation

Knowledge representation is a step where data visualization and knowledge representation tools are used to represent the mined data. Data is visualized in the form of reports, tables, etc.

| Sr. No. | Questions | Answer |
|---|---|---|
| 1 | Information can be converted into knowledge about ___ patterns and future trends. | Historical |
| 2 | Data about data is called ___. | Metadata |
| 3 | Facts, numbers, or text is called ___. | Facts, numbers, or text is called ___. |
| 4 | ___ and ___ are the key to emerging Business Intelligence technologies. | Data warehouse and data mining |
| 5 | Data mining is also called ___. | Knowledge discovery |
| 6 | Online Analytical Processing (OLAP) is a technology that is used to create ___ software. | Decision support |
| 7 | OLAP Supports ___ user access and multiple queries. | Multiple |
| 8 | Statistics techniques are incorporated into Data mining methods. (True/False). | True |
| 9 | ___ Optimization techniques are based on the concepts of genetic combination, mutation, and natural selection | Genetic algorithms |
| 10 | What is Mine set? | MineSet is software that provides tools for searching, sorting, filtering and drilling down enabling previously complex data models to be viewed intuitively through real-time 3-D graphical representation. |

| | | |
|---|---|---|
| 11 | A data warehouse refers to a database that is maintained separately from an organization's operational databases. (True/False) | True |
| 12 | A data warehouse is usually constructed by integrating multiple heterogeneous sources. (True/False) | True |
| 13 | ___ System is customer-oriented and is used for transaction and query processing by clerks, clients, and information technology professionals. | OLTP |
| 14 | A ___ allows data to be modeled and viewed in multiple Dimensions. | Data cube |
| 15 | . In ___ schema some dimension tables are normalized, thereby further splitting the data into additional tables. | Snowflake |
| 16 | ___ data model is commonly used in the design of relational databases. | Entity-relationship |
| 17 | Data warehouses and OLAP tools are based on ___ data model. | Multidimensional |
| 18 | The ___ exposes the information being captured, stored, and managed by operational systems. | Data source view |
| 19 | ___ are the intermediate servers that stand in between a relational back – end server and client front – end tools. | Relational OLAP (ROLAP) servers |
| 20 | A ___ is a set of views over operational databases. | Virtual warehouse |
| 21 | The ___ software gives the user the opportunity to look at the data from a variety of different dimensions. | Multidimensional Analysis |
| 22 | Which of the following statements defines Business Intelligence? | Converting data into knowledge and making it available throughout the organization and Analytical software and solutions for gathering, consolidating, analyzing and providing access |

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| | | |
|---|---|---|
| Lalpari Campus, Near Lalpari Lake, | | 3 – Vaishalinagar |
| B/H Marketing Yard, | | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | | Raiya Road, Rajkot |

| | | |
|---|---|---|
| | Ph. No. 9727753360 <br> Rajkot – 360001 <br> Ph.No.-(0281)2471645 | to information in a way that is supposed to let the users of an enterprise make better business decisions. |
| 23 | Based on the overall requirements of business intelligence, the ___ layer is required to extract, cleanse and transform data into load files for the information warehouse. | Data integration |
| 24 | Mining is not a business solution; it is just a technology. (True/False) | True |
| 25 | ___ is a random error or variance in measured variables. | Noise |

## Text & Web Analytics:

☐ Text analytics and text mining overview

- It is the procedure for generating the information and it will analyze the relation, pattern and the particular rule among all the database.

- Text mining is the process of exploring and analyzing large amounts of unstructured text data aided by software that can identify concepts, patterns, topics, keywords and other attributes in the data.

- It's also known as text analytics.

- It has 3 categories.

  - **Text Summarization**

  - **Text Categorization**

  - **Text Clustering**

**Text Summarization** : It is the procedure to extract its partial content which is reflected on the whole content.

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
|---|---|
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |

**Text Categorization** : We have to assign a special category to each of the text content which is predefined by the user.

**Text clustering** : We have to segment each text into several Cluster depends on the relevance.

- Text mining is a technology to discover patterns, trends and knowledge which is previously unknown, semi automatically from huge collections of unstructured text.
- Mining and analyzing text helps organizations find potentially valuable business insights in corporate documents, customer emails, call center logs, verbatim survey comments, social network posts, medical records and other sources of text-based data.

 Text mining applications

- We get benefits of text mining in our daily life often unknowingly. For example, the emails sent to us may be filtered through a text mining tool before being delivered to us.
- Like this there are many applications of text mining in various areas.
- Some applications of text mining are business, medicine, law, and society and so on.

How Text Mining is used in Business?

1. To know customer reactions (opinions)
2. To monitor competitor company's  Actions
3. To protect Intellectual Property

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

Lalpari Campus, Near Lalpari Lake,   3 – Vaishalinagar

B/H Marketing Yard,   Nr. Amrapali Railway Crossing

Amargadh (Bhichri), Rajkot   Raiya Road, Rajkot

Ph.No–(0281)2471645

4. To judge product's success and know Customer Satisfaction

5. Online recruitment

6. Helps in identifying Specific Knowledge Area

1. **To know customer reactions (opinions) :**

- This is done by different surveys like quantitative or descriptive surveys. In case of descriptive survey, questions accept natural language answers, which are very difficult to process.

- For this "Summarization" can help. In this sentiment analysis is used to categorize documents based on the nature of the text.

- Sentiment analysis identifies positive and negative opinions and emotions based on the occurrence of positive/negative words

2. **To monitor competitor company's Actions**

- With the help of "Information Extraction", Information monitor can scan sites of large companies to identify common articles, and

- summarize the information from multiple news sites in a single Web page.

- News articles are typically well written and somewhat easier to extract information from than a general Web page. Information about companies making investments, mergers, and related financial information can be captured in a template.

3. **To protect Intellectual Property**

- One way of protecting Intellectual Property is by filing patents. Big companies may file hundreds or even thousands of patents yearly.

- With thousands of patents on file, it is difficult to check the innumerable products launched yearly.

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
|---|---|
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |

Ph.No–(0281)2471645

- Manually matching thousands of products against thousands of patents is difficult and may cause some error.

- Text descriptions of products and patents can be automatically compared. Computations where the similarity is above a threshold can be considered for verification.

- For this document Clustering is useful.

4. **To judge product's success and know Customer Satisfaction**

- Customers feedback is taken in various forms like hard coded or through emails.

- These collected feedback needs to be classify to understand customers satisfaction level.

- Accurately categorizing email can be achieved with "Text Categorization/Classification".

5. **Online recruitment**

- Online recruitment is a huge industry. They collect

- resumes of candidate through emails or through a company's web site.

- A company may get hundreds or thousands of resumes.

- Handling these numbers of resumes is time consuming.

- An ideal resume of a candidate for a required vacancy with the flawless educational qualifications and work experience can be built.

- Attributes of submitted resumes and ideal resume can be compared to decide whether to call a candidate for further steps or not. This can reduce a lots of human efforts required against resume scrutiny to manage e-recruitment process.

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

Lalpari Campus, Near Lalpari Lake,                    3 – Vaishalinagar

B/H Marketing Yard,                    Nr. Amrapali Railway Crossing

Amargadh (Bhichri), Rajkot                    Raiya Road, Rajkot

Rajkot - 360001

## 6. Helps in identifying Specific Knowledge Area

- In a knowledge-rich society, there is too much information available for one person to absorb.

- When we want information on specialized topic, we need to consult experts for specialized expertise. Experts in an organization are usually known by the information published within an organization, recognition from external sources, and referral by examines.

- The flow conformation within an organization can be monitored, collected, and analysed to locate experts. This operation can occur transparently and does not require additional input from an employee. Information in the form of emails, presentations, Web pages, and formatted documents on an intranet provides a basis for identifying experts.

☐ Web mining overview

**Meaning:**

- Also known as WEB DATA MINING

- The term "**Web mining**" is coined by Etzioni (1996).

- It refers to discovery and analysis of useful information from the 'WORLD WIDE WEB'.

- Web Mining is the use of the data mining techniques to automatically discover and extract information from web documents/services.

**Goal:**

- Examine the use of data mining on the World Wide Web

- Using data mining techniques to make the web more useful and more profitable and to increase the efficiency of our interaction with the web

Web mining can be divided into three different types –

- **Web usage mining**

    **(Discovery of user access patterns from Web usage logs)**

- **Web content mining**

    **(To extract/mine useful information or knowledge from web page contents)**

- **Web structure mining**

    **(To discover useful knowledge from the structure of hyperlinks)**

| Sr. No. | Question | Answer |
|---------|----------|--------|
| 1 | ___ is used to examine the structure of a particular website and collate and analyze related data. | Structural mining |
| 2 | Which of the following techniques are concerned about user navigation accessing? | Web usage mining |
| 3 | Web data is ___. | Un-structured data |
| 4 | ___ Web mining involves the development of Sophisticated Artificial Intelligence systems. | an agent-based approach |
| 5 | The ___ approaches to Web mining have generally focused on techniques for integrating and organizing the heterogeneous and semi-structured data on the Web into more structured and high-level collections of resources. | Database |
| 6 | Association rules involving multimedia objects can be mined in ___ and ___ databases. | Image and video |
| 7 | In ___ approach, the signature of an image includes color histograms based on the color composition of an image regardless of its scale or orientation. | Color histogram-based signature |

| 8 | Which of the following are the measures of the text retrieval documents? | Precision, Recall and F-score |
| 9 | Data stored in most text databases are ___. | Semi-structured |
| 10 | Which of the following is the first step in text retrieval systems?. | Tokenization |
| 11 | Which of the following are the stop words? | A, The and of |
| 12 | Text databases are also called ___. | Document databases |
| 13 | Insurance and direct mail are two industries that rely on ___ to make profitable business decisions. | data analysis |
| 14 | To aid decision-making, analysts construct ___ models using warehouse data to predict the outcomes of a variety of decision alternatives. | Predictive |
| 15 | A ___ profile is a model that predicts the future purchasing behaviour of an individual customer, given historical transaction data for both the individual and for the larger population of all of a particular company's customers. | Predictive |
| 16 | Data mining can be used to help predict future patient behaviour and to improve treatment programs (True/False). | True |
| 17 | Data mining in the telecommunication industry helps to understand the business involved, identify telecommunication patterns (True/False). | True |
| 18 | GDP stands for ___. | gross domestic product |
| 19 | ___ is proving to be a critical link between theory, simulation, and experiment. | data-intensive computing |
| 20 | . IDS are based on ___ that are developed by the manual encoding of expert knowledge. | Handcrafted signatures |
| 21 | _____ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. | Text mining |
| 22 | In which database, data is a blend between structured and unstructured data formats? | Semi-structured data |
| 23 | The process of breaking out long-form text into sentences and words called? | Tokens |

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
|---|---|
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |

| 24 | Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences | True |
|---|---|---|
| 25 | Typical text mining tasks include? | text categorization, text clustering and entity relation modeling |
| 26 | Which of the following technique is not a part of flexible text matching? | Keyword Hashing |
| 27 | What is the right order for a text classification model components | . Text cleaning -> Text annotation -> Text to predictors -> Gradient descent -> Model tunin |
| 28 | What are the possible features of a text corpus? | Count of word in a document, Vector notation of word and Part of Speech Tag |
| 29 | What is the major difference between CRF (Conditional Random Field) and HMM (Hidden Markov Model) | CRF is Generative whereas HMM is Discriminative model |
| 30 | Stemming: This refers to the process of separating the prefixes and suffixes from words to derive the root word form and meaning. | True |
| 31 | The branch of statistics which deals with development of particular statistical methods is classified as | applied statistics |
| 32 | Which of the following is true about regression analysis? | modeling relationships within the data |
| 33 | Text Analytics, also referred to as Text Mining? | True |
| 34 | What is a hypothesis? | A statement that the researcher wants to test through the data collected in a study. |
| 35 | What is the cyclical process of collecting and analysing data during a single research study called? | Interim Analysis |
| 36 | The process of quantifying data is referred to as ___ | Enumeration |

| | Ph. No. 9727753360 | Rajkot - 360001 Ph.No. (0281)2471645 |
|---|---|---|
| 37 | An advantage of using computer programs for qualitative data is that they _ | Can reduce time required to analyse data (i.e., after the data are transcribed) Help in storing and organising data Make many procedures available that are rarely done by hand due to time constraints |
| 38 | Boolean operators are words that are used to create logical combinations.<br><br>True | True |
| 39 | _____ are the basic building blocks of qualitative data.<br><br>Categories | Categories |
| 40 | This is the process of transforming qualitative research data from written interviews or field notes into typed text.<br><br>Transcription | Transcription |
| 41 | A challenge of qualitative data analysis is that it often includes data that are unwieldy and complex; it is a major challenge to make sense of the large pool of data.<br><br>True | True |
| 42 | Hypothesis testing and estimation are both types of descriptive statistics.<br><br>False | False |
| 43 | A set of data organised in a participants(rows)-by-variables(columns) format is known as a "data set." | True |

| 44 | A graph that uses vertical bars to represent data is called a ___ | Bar graph |
|----|-----|-----|
| 45 | ____ are used when you want to visually examine the relationship between two quantitative variables. | Scatter plot |
| 46 | The denominator (bottom) of the z-score formula is what? | The standard deviation |
| 47 | Which of these distributions is used for a testing hypothesis? | Chi-Squared Distribution |
| 48 | A statement made about a population for testing purpose is called? | Hypothesis |
| 49 | If the assumed hypothesis is tested for rejection considering it to be true is called? | Null hypothesis |
| 50 | If the null hypothesis is false then which of the following is accepted? | Alternative Hypothesis. |
| 51 | Alternative Hypothesis is also called as? | Research Hypothesis |
| 52 | Quantitative content analysis is an approach that aims to: | a) Objectively and systematically measure the content of a text |

☐ Social media analytics

- Social Media is a Web and mobile based Internet applications that allow the creation, access and exchange of user generated contents

- A powerful tool to change the world order

- Majority of them yield unstructured data

- "Social media analytics is the practice of gathering data from blogs and social media websites, such as Twitter, Facebook, etc and analysing that data to inform business decisions.

- Social Media Analytics deals with development and evaluation of tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data

The most common use of social media analytics is gauging customer opinion to support marketing and customer service activities.



Importance of Social Media Analytics

- Social media analysis is a valuable  tool that can help you meet your most important and strategic business goals.

**1. Social media analytics helps organization to understand their targeted audience.**

2. S**ocial media analytics can increase engagement and responsiveness**:

3.  **Social media analytics can highlight problems and weaknesses to discover new trends and avoid a brand crisis.**

**4. Social media analytics can help you learn from the competition**

Social Media Measurement Tool

Lalpari Campus, Near Lalpari Lake,                          3 – Vaishalinagar

B/H Marketing Yard,                          Nr. Amrapali Railway Crossing
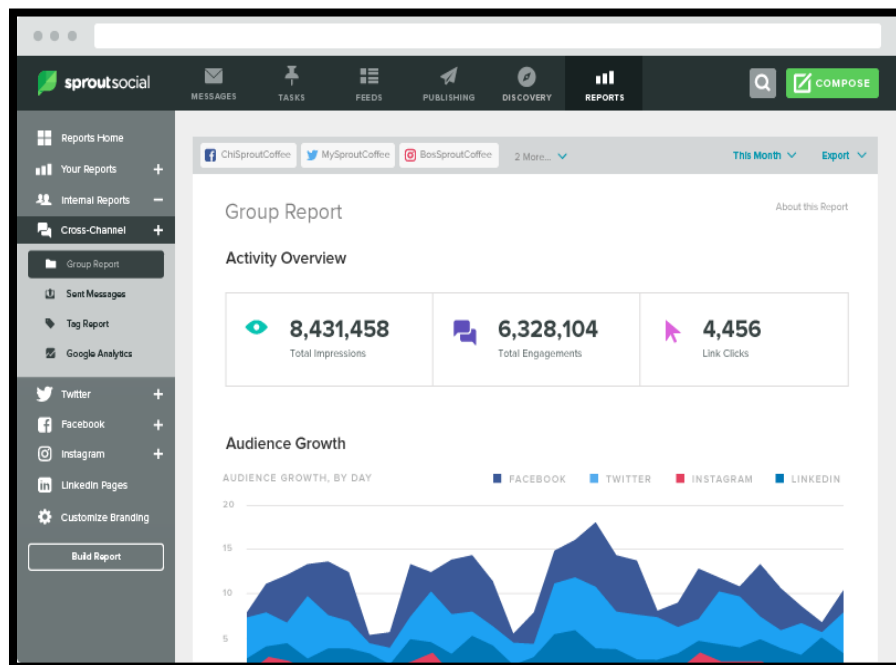
Amargadh (Bhichri), Rajkot                          Raiya Road, Rajkot

# 1. Sprout social :    Ph. No. 9727753360                          Rajkot - 360001

- you can measure performance across Facebook, Twitter, Instagram and LinkedIn, all within a single platform.

- Having all of your analytics in one place makes it easier to track and compare your efforts across multiple profiles and platforms.

  – **Networks:** Facebook, Twitter, Instagram, LinkedIn and Google+
  – **Price:** Starts at $99/month (**try a free 30-day trial**)

**Recommended for:** Any brand that manages multiple social media profiles across multiple networks. If your brand is active on social media, a tool like Sprout is a must-have.



# 2. Iconosquare :

- Iconosquare is a social media analytics tool specifically for Instagram.

- One of the standout features that separates Iconosquare from other tools is that in addition to analysis of your normal photos and videos. It also gives you insights into Instagram stories. With higher level plans, you can also get influencer analytics as well.

  – Networks: Instagram

  – Price: Starts at $9/month

  – Recommended for: Brands heavily invested into Instagram marketing.
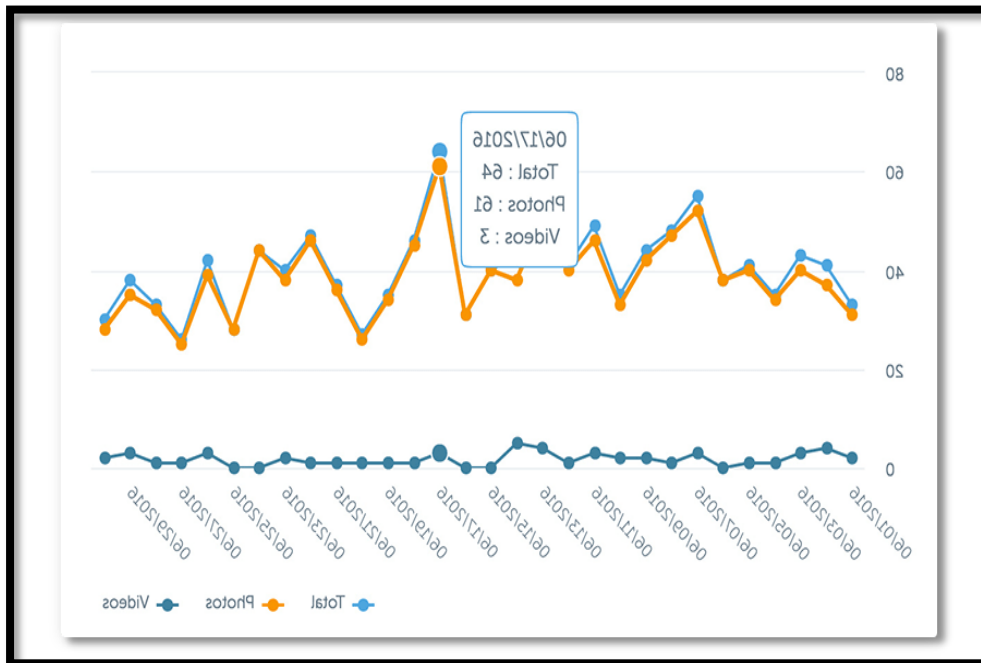


## 3. Tailwind

- While Instagram and Snapchat are currently the most talked about players in the visual social media landscape, Pinterest is still very active. And just like with any other social network, you need to measure your

performance. **Tailwind** is arguably the most popular third-party Pinterest analytics tool.

- Through Tailwind, you can track trends in followers and engagement, analyze your audience and they even provide some **Instagram analytics** as well at certain plan levels.

  - **Networks:** Pinterest and Instagram
  - **Price:** Starts at $9.99/month

**Recommended for:** Brands that use Pinterest as one of their top marketing channels
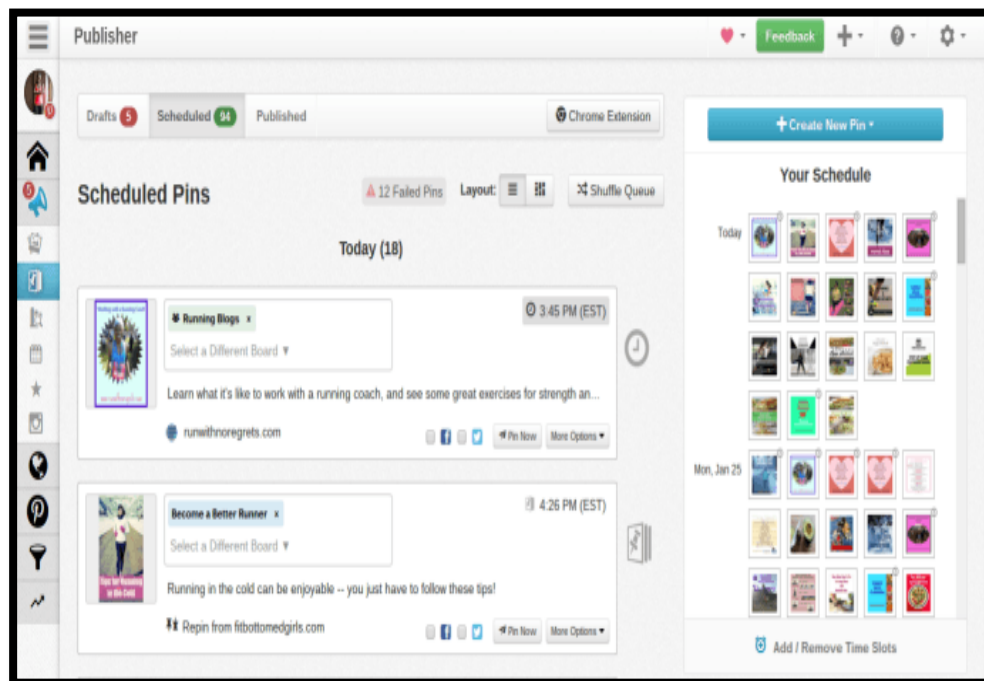


## 4. Google Analytics

- While it's not technically a "social media analytics tool," Google Analytics (GA) is one of the best ways to track social media campaigns and even help you **measure social ROI**.

- You likely already have GA setup on your website to monitor and analyze your traffic. But did you know you can access and create reports specifically for social media tracking?

- For instance, you can see how much traffic comes to your website from each social network, or **use UTM (Urchin Tracking Module) parameters** to **track specific social media campaigns**. **UTM**

  – **Networks:** All

  – **Price:** Free

  – **Recommended for:** Any brand with a website.

| | Social Network | Acquisition | | | Behavior | | | Conversions | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sessions | % New Sessions | New Users | Bounce Rate | Pages / Session | Avg. Session Duration | Goal Conversion Rate | Goal Completions |
| | | 180 <br> % of Total: 0.55% (32,509) | 86.11% <br> Avg for View: 90.85% (-5.22%) | 155 <br> % of Total: 0.52% (29,536) | 92.78% <br> Avg for View: 91.13% (1.81%) | 1.12 <br> Avg for View: 1.14 (-1.43%) | 00:00:23 <br> Avg for View: 00:00:39 (-41.16%) | 0.00% <br> Avg for View: 0.00% (0.00%) | 0 <br> % of Total: 0.00% (0) |
| ☐ | 1. Facebook | 140 (77.78%) | 85.71% | 120 (77.42%) | 92.86% | 1.12 | 00:00:17 | 0.00% | 0 (0.00%) |
| ☐ | 2. Pinterest | 22 (12.22%) | 86.36% | 19 (12.26%) | 95.45% | 1.14 | 00:00:15 | 0.00% | 0 (0.00%) |
| ☐ | 3. Twitter | 9 (5.00%) | 88.89% | 8 (5.16%) | 77.78% | 1.22 | 00:02:27 | 0.00% | 0 (0.00%) |
| ☐ | 4. StumbleUpon | 6 (3.33%) | 100.00% | 6 (3.87%) | 100.00% | 1.00 | 00:00:00 | 0.00% | 0 (0.00%) |
| ☐ | 5. Blogger | 1 (0.56%) | 0.00% | 0 (0.00%) | 100.00% | 1.00 | 00:00:00 | 0.00% | 0 (0.00%) |
| ☐ | 6. Instagram | 1 (0.56%) | 100.00% | 1 (0.65%) | 100.00% | 1.00 | 00:00:00 | 0.00% | 0 (0.00%) |
| ☐ | 7. LinkedIn | 1 (0.56%) | 100.00% | 1 (0.65%) | 100.00% | 1.00 | 00:00:00 | 0.00% | 0 (0.00%) |

☐ Sentiment analysis overview

- The process of analyzing unstructured text to extract relevant information and transforming it into useful business intelligence

- It determines if an expression is positive, negative, or neutral, and to what degree

It is an emerging field that attempts to analyze and measure human emotions and convert it into hard facts

- **Also called opinion mining since it includes identifying attitudes, emotions, and opinions of a company's product, brand, or service**
- **It is text analytics that looks at the face value of the words to give them meaning**
- **Gives insight into the emotion behind the words**
- **Helps businesses monitor news articles, online forums and social networking sites for trends in opinions about their products and services**
- **Businesses have realized that not all opinions are equally important-Some opinions carry more weight than others**
  - **A negative tweet by Lady Gaga will have a much greater impact than a tweet by an ordinary person**
- **It is a tool to allow users to generate 'influence scores' to identify people, blogs, forums etc. that are important**

**How does it Work?**

**A look at the different processes involved with Sentiment Analysis**

**Data Collection**

- **Public sentiments from consumers expressed on public forums and on social network are collected**

- **Opinions and feelings are expressed in different way, with different vocabulary, context of writing, usage of short forms and slang, makes data huge and disorganized**

## Analyse Data

- **Text Preparation**
- **Data is extracted and filtered before analysis**
- **Non-textual content and content is identified and eliminated if it is irrelevant**
- **Each sentence and opinion is examined for subjectivity**
- **Sentences with subjective expressions are retained and ones that convey objective expressions are discarded**

## Indexing

- **Sentiments can be broadly classified into two groups, positive and negative**
- **Each subjective sentence is classified into positive, negative, good, bad, like, dislike**

## Delivery

- **(Presentation of Output)**
- **The result of converted unstructured text into meaningful information**
- **Usually displayed as graphs for easy interpretation**

## Sentiment analysis

**Big Data Analytics:**

☐ Definition and characteristics of big data

- Big Data is a complex sets of Data. Big Data is a collection of huge volumes of Data.

- Generally Big Data consists unstructured Data

- Bid Data is the Data characterized by 4 key attributes/Characters :

➢      **Volume**  (Vast Amount of data generated at every second)

➢      **Variety**   (Different kind of Data)

➢      **Velocity**  ( Speed at which new data is generated and moved)

➢      **Value**     (Trustworthiness of Data)

- The basic idea behind the phrase 'Big Data' is that everything we do is increasingly leaving a digital data, which we can use and analyze.

- Big Data therefore refers to our ability to make use of the every increasing volumes of data.

- Byte                    One Grain of Rice

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

Lalpari Campus, Near Lalpari Lake,                3 – Vaishalinagar

B/H Marketing Yard,                    Nr. Amrapali Railway Crossing

Amargadh (Bhichri), Rajkot              Raiya Road, Rajkot

Ph. No. 07755460661          Rajkot - 360001

Ph.No–(0281)2471645

- KiloByte          Cup of Rice

- MegaByte          8 Bags of Rice

- GigaByte          3 Trucks

- TeraByte          2 Container Ships

- PetaByte          Fills the districts

- ExaByte          Fills West Coast States

- ZettaByte          Fills the Pacific Ocean

- YottaByte          An Earth Size Rice Ball..!

- Byte          One Grain of Rice          Mini Computer (Old)

- KiloByte          Cup of Rice

Example

- MegaByte     8 Bags of Rice          Desktop

- GigaByte     3 Trucks

- TeraByte     2 Container Ships          Internet

- PetaByte     Fills the districts

- ExaByte     Fills West Coast States   Big Data- facebook, amazon, e-bay, google

- ZettaByte     Fills the Pacific Ocean

- YottaByte     An Earth Size Rice Ball..!          The Future..!

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| | |
|---|---|
| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |
| Ph. No. 9727753360 | Rajkot - 360001 |
| | Ph.No–(0281)2471645 |

☐ Fundamentals of big data analytics

- It's widely accepted today that the phrase "big data" implies more than just storing more data.

- It also means doing more with data.

- Big data analytics or data science probably come closest. We can probably refine the various techniques into three big groups:

➢ **Predictive analytics**

➢ **Collective intelligence,**

➢ **Machine learning**

**Predictive analytics**

➢ It is a class of algorithms that use data from past to predict future.

➢ With the help of mathematical function like regression analysis,linear regression,time series analysis predicting the future.

**Collective intelligence**

- Collective intelligence sounds like a complex academic pursuit, but it's actually something we encounter every day.

- When Google or another search engine corrects or predicts your searches, it is using the data collected from the billions of other peoples' searches that came before yours.

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

Lalpari Campus, Near Lalpari Lake,                    3 – Vaishalinagar

B/H Marketing Yard,                                    Nr. Amrapali Railway Crossing

Amargadh (Bhichri), Rajkot                             Raiya Road, Rajkot

Ph. No. 9727753360                                     Rajkot - 360001

Ph.No–(0281)2471645

## Machine learning

- Machine learning is a field of computer science.

- A branch of Artificial Intelligence.

- Getting computers to program themselves

- Machine Learning is concerned with computer programs that automatically improve their performance through experience

## Types of Machine Learning

1. A **Supervised Machine Learning** algorithm is one that requires some training in order to build a model.

   – The majority of practical machine learning uses supervised learning.

   – It is a type of system in which both input and desired output data are provided

2. An **Unsupervised machine learning** requires no training sets

   – A type of Machine Learning Algorithm used to draw inferences from datasets consisting of input data without labeled responses (output).

   – Self Learning
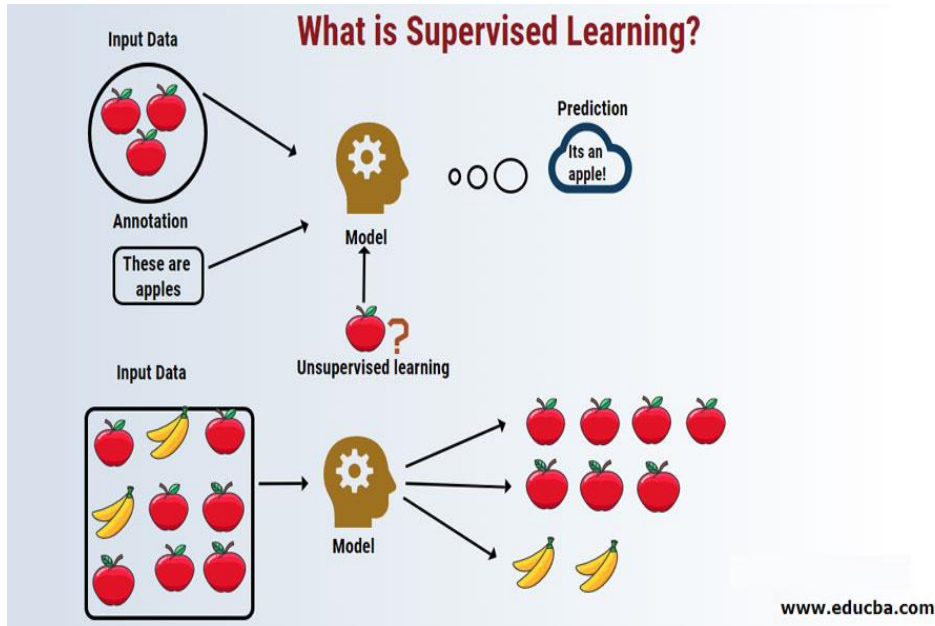
Lalpari Campus, Near Lalpari Lake,                    3 – Vaishalinagar

B/H Marketing Yard,                    Nr. Amrapali Railway Crossing

Amargadh (Bhichri), Rajkot                    Raiya Road, Rajkot

# Application Of Big Data analytics

Smarter Healthcare

Homeland Security

Traffic Control

Manufacturing

Multi-channel sales

Telecom

Trading Analytics

Search Quality

| Sr. No. | Questions | Answer |
|---|---|---|
| 1 | In order to identify the users, web analytics tools need to report on? | User session |
| 2 | The most common user identification technique is via? | Cookies |
| 3 | The focus of web analytics is to understand the? | users of a site, User behavior and User activities |
| 4 | The immediate purpose of analyzing digital analytics data is to: | Make better decisions about your business |
| 5 | Which of the following statements is incorrect? | You should provide usability-related (heuristics) recommendations without actual data in your dashboards. |

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

| Lalpari Campus, Near Lalpari Lake, | 3 – Vaishalinagar |
| B/H Marketing Yard, | Nr. Amrapali Railway Crossing |
| Amargadh (Bhichri), Rajkot | Raiya Road, Rajkot |

| 6 | What are the three areas of expertise of a Center of Excellence? | Business Outcomes and Actions / Enabling Capabilities / Analysis |
|---|---|---|
| 7 | You want to understand if users use their mobile phone to access your site. Which of the following approach is preferable? | Look at the report Audience/Mobile/Overview |
| 8 | Publishing ads through Facebook would fall under which media type? | Paid media |
| 9 | Google Analytics can only recognize returning users on websites, not on mobile apps. | False |
| 10 | If I want to have an estimate of the number of people who visited my website, which metric should I use? | Unique visitors |
| 11 | Above all, the advantage of competitive intelligence is that it allows you to | Measure your website data compared to your competitors |
| 12 | What would be a good way to segment your data?<br><br>A. Segment your primary geographic market<br>B. Segment based on Acquisition / Behavior / Conversion dimensions<br>C. Segment based on Recency, Frequency, and | Monetary Value |
| 13 | Which of these marketing efforts would NOT be an effective use of campaign trackers? | Television advertising campaigns |
| 14 | Which of the following reports would identify the entry pages that are contributing the most to your e-commerce revenues? | Landing Pages |
| 15 | What report identifies browsers that may have had problems with a website? | The Browser & OS report |
| 16 | What report shows which web pages get the most traffic and highest engagement? | The All Pages report |
| 17 | What feature is required to collect the number of comments users posted to a web page? | Custom Metric |
| 18 | When it comes to web analytics, what insights can you gather using analytics tools? | How people interact with your website |
| 19 | If a web page visitor clears the Analytics cookie from their browser, what will occur? | Analytics will not be able to associate user behavior data with past data collected |

**SHREE H.N.SHUKLA COLLEGE OF MANAGEMENT STUDIES, RAJKOT**
**AFFILIATED TO GUJARAT TECHNOLOGICAL UNIVERSITY**

Lalpari Campus, Near Lalpari Lake,       3 – Vaishalinagar

B/H Marketing Yard,       Nr. Amrapali Railway Crossing

Amargadh (Bhichri), Rajkot       Raiya Road, Rajkot

Ph. No. 9727753360       Ph.No–(0281)2471645

| | | |
|---|---|---|
| | | Analytics will set a new unique ID the next time a browser loads a tracked page<br>Analytics will set a new browser cookie the next time a browser loads a tracked page |
| 20 | Which ONE of the following is based on user-generated media, mainly investigating earned media? | Social media monitoring |
| 21 | The objectives for web analytics are likely to concern: | Measurement of web site performance |
| 22 | Which of these extracts information from user generated content for further scrutiny? | Web-scraping software |
| 23 | Which of the following is the odd one out? | Share of conversation |
| 24 | Which ONE of the following is mainly used in Web Analytics and is free of charge? | Google Analytics |
| 25 | Social networks are organized primarily around _____. | People |
| 26 | Which social network is considered the most popular for social media marketing? | Facebook |
| 27 | What is the name for Facebook`s ranking algorithm? | Edge rank |
| 28 | Which of the following is an important aspect of creating blogs and posting content? | Using a witty user name , Posting at least once a month to the blog and Social Media Optimization |
| 29 | What is meant by "micro-blogging"? | Blogs with limited individual posts, limited by character count typically |
| 30 | What is "social media optimization"? | Creating content which easily creates publicity via social networks |
| 31 | What would the marketing budget section of a marketing plan detail? | The expected costs for each ad campaign based on the delivery method |
| 32 | What is the name of Facebook's analytic package? | Insights |

| 33 | How does creating a social network marketing plan differ from a traditional marketing plan? | The staff requirements and skill sets for social marketing are different |
|----|---|---|
| 34 | Which of the following is functions of social media for business? | Are you participating in the conversation and sharing? And Are you listening and monitoring what is being said about you? |
| 35 | What feature does LinkedIn offer for pay accounts? | Increased abilities to connect directly and send messages to people |
| 36 | Why is it important to post to a blog regularly? | Keep readers engaged and also gives search engines content to index |
| 37 | Why is it important to post to a blog regularly? | Keep readers engaged and also gives search engines content to index |
| 38 | How can a company use the same material for both traditional and social network marketing? | Utilizing a television ad campaign online as well on their site and sites such as YouTube |
| 39 | How is site traffic useful in evaluating marketing? | Ads can send receivers to a specific landing page, which can be tracked |
| 40 | How does a blog directly impact sales of a company? | It is typically used as the way to disperse company coupons |
| 41 | What traditional marketing technique is YouTube closest to? | Television advertising |
| 42 | What is another term for "social media"? | Consumer Generated Media |
| 43 | What is meant by "Marketing Creative"? | The content for marketing and its creative aspect |
| 44 | What is one measure a company can use to validate the usefulness of its video posts on YouTube? | The amount of views of the video |
| 45 | How can a company convert posts on Twitter to sales? | Creating posts which drive followers to their site |
| 46 | Which of the following method used to identify user sessions? | IP address, authenticated user & user agent |